



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

[Signature]

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/621,474	07/17/2003	Hiroshi Nomiyama	JP920020109US1	2462
25259	7590	02/22/2007	EXAMINER	
IBM CORPORATION 3039 CORNWALLIS RD. DEPT. T81 / B503, PO BOX 12195 REASEARCH TRIANGLE PARK, NC 27709			ROSE, HELENE ROBERTA	
		ART UNIT	PAPER NUMBER	
		2163		

SHORTENED STATUTORY PERIOD OF RESPONSE	NOTIFICATION DATE	DELIVERY MODE
3 MONTHS	02/22/2007	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

If NO period for reply is specified above, the maximum statutory period will apply and will expire 6 MONTHS from the mailing date of this communication.

Notice of this Office communication was sent electronically on the above-indicated "Notification Date" and has a shortened statutory period for reply of 3 MONTHS from 02/22/2007.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

RSWIPLAW@us.ibm.com

Office Action Summary	Application No.	Applicant(s)
	10/621,474	NOMIYAMA ET AL.
	Examiner	Art Unit
	Helene Rose	2163

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) Responsive to communication(s) filed on 11 December 2006.
- 2a) This action is FINAL. 2b) This action is non-final.
- 3) Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) Claim(s) 15-17 and 21-24 is/are pending in the application.
- 4a) Of the above claim(s) 1-14 and 18-20 is/are withdrawn from consideration.
- 5) Claim(s) _____ is/are allowed.
- 6) Claim(s) 15-17 and 21-24 is/are rejected.
- 7) Claim(s) _____ is/are objected to.
- 8) Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) The specification is objected to by the Examiner.
- 10) The drawing(s) filed on 17 July 2003 is/are: a) accepted or b) objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) All b) Some * c) None of:
 1. Certified copies of the priority documents have been received.
 2. Certified copies of the priority documents have been received in Application No. _____.
 3. Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- | | |
|--|---|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892) | 4) <input type="checkbox"/> Interview Summary (PTO-413) |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948) | Paper No(s)/Mail Date. _____ |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO/SB/08) | 5) <input type="checkbox"/> Notice of Informal Patent Application |
| Paper No(s)/Mail Date _____ | 6) <input type="checkbox"/> Other: _____ |

Detailed Action

1. In response to communications entered on 11/29/2006, Claims 15-17 were amended; Claims 8, 10-12, and 18-20 were cancelled; Claims 21-24 were added.
2. Applicant's arguments entered on 11/29/2006, with respect to the rejected claims in view of the cited references have been considered but are moot in view of applicant's amended claims necessitate new ground(s) of rejection.

Claim Rejections – 35 U.S.C 103

3. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.
4. Claims 15-17 and 21-24 are rejected under 35 U.S.C. 103(a) as being unpatentable over Mantha et al. (US Patent No. 6,163,779, Date of Patent: December 19, 2000, hereinafter Mantha) in view of Wyler (US Patent No. 7,047,033, Date Filed: January 31, 2001).

Claims 15 and 21:

Regarding claims 15 and 21, Mantha teaches reading an HTML document of a web page as an analyzing object (Figure 14, all features, wherein it's further defined in column 12, lines 42-47, wherein the upper portion of the code represents the original HTML source code, wherein in lower portion each if the href tags has been modified to

point to the local storage, which is interpreted to be equivalent to “reading the source codes of said web pages from said storage device”, as described in paragraph [0018], of applicant specification, Figure 21, Mantha);

conducting a temporary block analysis based on a description of HTML tags of the HTML document (column 2, lines 29-38, wherein the original page, i.e. the base HTML document, is then parsed to prepare a list of hypertext references, wherein this is interpreted to be “conducting a block analysis”, wherein such references are typically represented by <a href> markup tags, wherein for each reference tag in the base HTML document, wherein this is interpreted to be “description of HTML tags of the HTML document”, that is an embedded object, e.g., an image, a copy of that file is retrieved from the server and then saved on the local hard drive, wherein the new HTML page, the path name to the stored file is substituted for the original hypertext reference, wherein “substitute” is interpreted to be equivalent to “temporary”, and therefore, interpreted to be equivalent to “conducting a temporary block analysis based on a description of HTML tags of the HTML document”, Mantha);

using the HTML tags to temporarily divide the HTML document into blocks (column 2, lines 28-30, wherein the original page, i.e., the base HTML document, is then parsed to prepare a list of hypertext references, wherein the term “parsed” is interpreted to be a method of “dividing”, and lines 37-38, wherein the path name to the stored file is substituted for the original hypertext references, which is interpreted to be equivalent to “using the HTML tags to temporarily divide the HTML document into blocks”, Mantha);

Art Unit: 2163

Mantha discloses the following limitations above. However, Mantha does not disclose wherein “identifying unnecessary information elements in the HTML document”.

On the other hand, Wyler discloses identifying unnecessary information elements in the HTML document (column 12, lines 4-8, wherein the application removes irrelevant information, such as images and data, i.e., advertising banners, links to unrelated issues, from the webpage, wherein a web page is a document written in Hypertext Markup Language, which is interpreted to be equivalent to an “HTML” document, Wyler).

It would have been obvious to one of the ordinary skill in the art at the time of the invention to incorporate Wyler teachings into Mantha system. A skilled artisan would have been motivated to combine as suggested by Wyler [column 12, lines 1-2], in order to recognize irrelevant data. As a result, establishing an improved method of providing tailored information and a suitable display to a user.

wherein the unnecessary information elements are plural information elements that include:

plural information elements that include an OBJECT_IMAGE having a same Uniform Resource Locator (URL), wherein the OBJECT_IMAGE describes a type of media used to display the HTML document (column 22, lines 20-28, wherein Banner Advertisements, in which advertisement objects that appear in the document in the form of banners, which may include image and/or links, and Image Advertisements wherein these images that appear in the HTML page with no relevance to the page subject; and column 22, line 57, wherein this is a background image of the web/HTML page, wherein background image is interpreted to be the static image that appears behind text, graphics, and other web page components, which is interpreted to be equivalent to

“plural information elements that include an OBJECT_IMAGE having a same Uniform Resource Locator (URL), wherein the OBJECT_IMAGE describes a type of media used to display the HTML document”, Wyler),

text in the HTML document that is shorter than a maximum predetermined length (column 31 and 32, lines 65-67 and line 7, wherein the logical location of an object which is interiorly disposed relative to the base object is the maximum value of 100, wherein the logical location of any other object is the distance on the webpage of that object from the base object, and wherein distance is equivalent to “length”, Wyler)
and wherein the text appears in the HTML document more than a predetermined frequency (column 32, lines 47-52, wherein base object is selected which is the largest object on the webpage, wherein “web page” is interpreted to be equivalent to a HTML document, if there is a tie, i.e. if the largest two or more objects are similar to a predetermined extent in size, “which is interpreted to be equivalent to a “predetermined frequency”, then the object with the most words, wherein this is equivalent to “text appears in the HTML document”, in its typically deemed to the base object, in which is interpreted to be equivalent to “wherein the text appears in the HTML document more than a predetermined frequency, Wyler),

multiple anchors having a same title (column 11, lines 8-13, wherein identifying the URL’s in the original file and storing them, and wherein duplicates entries are eliminated, and wherein URL’s are defined to be An Internet World Wide Web Address, which is interpreted to be equivalent to ‘anchors’, in which anchors are the same as hyperlinks, and wherein “duplicate” is defined to be “to reproduce a page or an image exactly the same”, which is interpreted to be equivalent to “same title”, Mantha),

image tags that only perform a role of punctuation for text in the HTML

document (column 11, lines 43-61, wherein the application searches the web page source or an input text file from markup languages, which is equivalent to a “HTML document”, and wherein the application passes the page content to one of the three following functions, which is interpreted to be equivalent to a “role”, and wherein the markup language parses and analyze the markup languages, and wherein the rich text format parses and analyze the text by taking common knowledge of the text format, like bigger font size, which is interpreted to be equivalent to “image tags that only perform a role of punctuation for text in the HTML document”, Wyler), **and**

multiple text blocks having a same description (Figure 9, diagrams 61-62, 64, and 66, wherein further define in column 9, lines 50-67, wherein a representative web page includes a base HTML document 62, and one or more embedded objects such as .gif file and jpeg files, wherein the .gif and jpeg files are interpreted to be the “text blocks” which is interpreted to be equivalent to “illustrations”, and wherein the copy operating saves one HTML page corresponding to the base HTML document and one file, e.g. reference numerals for each saved embedded object, wherein the new HTML page is the local copy of the original file, wherein typically the two pages are the same, which is interpreted to be equivalent to “multiple text blocks having the same description, Mantha);

defining any block in the HTML document that is deemed to be meaningless as **an OBJECT**, wherein a block is deemed to be meaningless if that block **contains** only unnecessary information elements **and at least one anchor** (column 12, lines 4-8, wherein the application removes irrelevant information from the webpage and

Art Unit: 2163

reorganizes the information into objects with categories in a file represent by the M2O script language, which is interpreted to be equivalent to “defining any blocks in the HTML document that is deemed to be meaningless as an OBJECT DELIMINTER, and wherein irrelevant information includes links to unrelated issues, advertising banners, and images, which is interpreted to be” wherein a block is deemed to be meaningless if that block contains only unnecessary information elements and at least one anchor”, Wyler);

crawling only anchors found in blocks that have not been defined as OBJECT DELIMITER’s (column 14, lines 39-44, wherein application searches all the documents for words that fit into the index category and when finding such words the application program inserts an index command, in which from that point on the web pages are called documents, which is interpreted to be equivalent to “crawling only anchors found in blocks that have not been defined as OBJECT DELIMITER’s”, Wyler).

Claims 16 and 22:

Regarding claims 16 and 22, teaches wherein the **maximum predetermined length is 12 bytes** (column 11, lines 27-28, wherein the object which is the biggest or has the most number of words in it; column 13, lines 58-61, wherein object size is a value equal to width * height of the object; column 32, line 50, wherein predetermined “extent” is interpreted to be equivalent to “length”, Wyler).

Claims 17 and 23:

Regarding claims 17 and 23, teaches wherein the **predetermined frequency is ten times** (column 16, lines 39-41, wherein the mechanism for selecting the relevant objects is based on selecting the objects with weights that pass the predefined thresholds, which is interpreted to be equivalent to “predetermined frequency”; and column 19, lines

Art Unit: 2163

3-7 and 33-36, wherein the application tries to reduce the index list length by finding identical words with different page numbers and wherein if chapters and sections with titles and subtitles do not appear in the same document after the third level only the following changes takes place, which is irrelevant images/data are taken off; column 31, lines 50, wherein system counts the number of words in the object which do not occur in the base object, wherein the proportion of words in the object which occur in the base object from among the total number of words in the object, which determines the word matching, which is equivalent to frequency, Wyler).

Claim 24:

Regarding Claim 24, SEE claims 15 and 21 above, wherein this limitation is substantially the same/or similar).

Prior Art of Record

1. Wyler et al (US Patent No. 7,047,033)
2. Ishikawa et al (US Patent No. 5,848,407)
3. Finseth et al (US Patent No. 6,271,840)
4. Mantha et al (US Patent No. 6,163,779)
5. Wang Baldonado (US Patent No. 6,704,722)

Art Unit: 2163

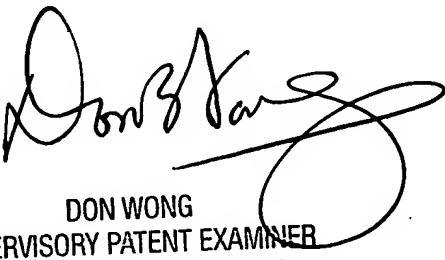
Point of Contact

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Helene Rose whose telephone number is (571) 272-0749. The examiner can normally be reached on 8:00am - 4:30pm Monday-Friday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Don Wong can be reached on (571) 272-1834. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

HRR
Technology Center 2100
February 5, 2007



DON WONG
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 2100